Improvements

1. CommunityPulse.py
   * Removed hardcoded secrets from Reddit API config for better security.
   * Added logging for database connections and errors.
   * Added Streamlit spinners for scraping and cached expensive operations (data load, insights).
   * Added input/data validation before visualizations to prevent errors and improve UX.
   * Improved comments and code clarity.
2. requirements.txt
   * Cleaned up and consolidated dependencies.
   * Removed unused packages (gymnasium, torchvision, torchaudio, beautifulsoup4).
   * Pinned all package versions for reproducibility.
   * Added a clear note to install the spaCy model after requirements install.
3. CommPulseRequirements.txt
   * Marked as deprecated and now points users to requirements.txt.

Next steps:

* Test the app (including scraping, dashboard interactions, and visualizations).
* Make sure to set your Reddit/MongoDB credentials as environment variables.
* After installing requirements, run:
* “python -m spacy download en\_core\_web\_sm” in Terminal

Summary of Actions:

* CommunityPulse.py: Improved security, error handling, UX, and maintainability.
* requirements.txt: Cleaned up, pinned, and clarified dependencies.
* CommPulseRequirements.txt: Marked as deprecated and redirected users to requirements.txt.

Next Steps:

1. Use only requirements.txt to install dependencies:
2. Insert in Terminal
3. pip install -r requirements.txt
4. After installing, run:
5. Insert in Terminal
6. python -m spacy download en\_core\_web\_sm
7. Set your Reddit and MongoDB credentials as environment variables (do not hardcode).
8. Test your application to ensure all features work as expected.

## etup Instructions

1. Clone the repository
2. bash
3. CopyInsert

git clone <your-repo-url>

1. cd nlpproject1/RL\_StreamlitApp
2. Install dependencies
3. bash
4. CopyInsert in Terminal
5. pip install -r requirements.txt
6. Download spaCy English model
7. bash
8. CopyInsert in Terminal
9. python -m spacy download en\_core\_web\_sm
10. Set up environment variables Create a .env file or set these variables in your environment:
11. CopyInsert

REDDIT\_CLIENT\_ID=your\_reddit\_client\_id

REDDIT\_CLIENT\_SECRET=your\_reddit\_client\_secret

REDDIT\_USER\_AGENT=your\_custom\_user\_agent

1. MONGODB\_URI=your\_mongodb\_connection\_string
2. Run the Streamlit app
3. bash
4. CopyInsert in Terminal
5. streamlit run CommunityPulse.py

## Usage

* Use the dashboard sidebar to scrape new data, select time filters, and explore insights.
* Visualizations update automatically based on the latest data.
* All data is stored in MongoDB for persistence and scalability.

## Contributors

* Jesse Katz
* Hung Lu

## License

This project is for educational purposes.